



## Recherche de patterns dans un corpus d'actions multimodales

Marie-Laure Betbeder, Régis Tissot, Christophe Reffay

### ► To cite this version:

Marie-Laure Betbeder, Régis Tissot, Christophe Reffay. Recherche de patterns dans un corpus d'actions multimodales. Environnement Informatique pour l'Apprentissage Humain, Jun 2007, Lausanne, Suisse. pp.533-544. edutice-00158881

**HAL Id: edutice-00158881**

**<https://edutice.archives-ouvertes.fr/edutice-00158881>**

Submitted on 1 Jul 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Recherche de patterns dans un corpus d'actions multimodales

**Marie-Laure Betbeder, Régis Tissot, Christophe Reffay**

*LIFC – Université de Franche-Comté  
16 Route de Gray  
25030 Besançon Cedex  
{prenom.nom}@lifc.univ-fcomte.fr*

---

*RÉSUMÉ. Nous présentons dans cet article nos travaux sur la recherche de patterns dans un corpus d'actions multimodales issues d'une formation en ligne. De par la multitude et l'hétérogénéité des traces issues d'une formation en ligne, il devient difficile d'avoir une bonne appréhension des interactions. L'outil proposé a pour objectif d'aider un chercheur à analyser des données en repérant les épisodes récurrents (patterns). Ce premier repérage automatique peut ensuite être étudié plus précisément par l'analyste. Nous présentons ici, un premier bilan de l'outil implémenté.*

*MOTS-CLÉS : recherche de patterns, corpus, interactions multimodales.*

---

## 1. Introduction - Contexte

Les formations se déroulant dans des dispositifs d'apprentissage en ligne génèrent une multitude de traces de format et de type différents. Lorsqu'il s'agit de dispositifs synchrones, les traces sont encore plus nombreuses et leur nature implique un « enchevêtrement » et/ou chevauchement des interactions. Avant de pouvoir analyser les interactions et tenter de comprendre les phénomènes qui se sont déroulés, il est indispensable de travailler sur « l'organisation, la modélisation et la conceptualisation des traces d'activité, de leur représentation et de leur traitement » [SETTOUTI et al.]. Dans [BETBEDER et al. 06] nous avons proposé une organisation et une modélisation des données issues de ce genre de dispositif. Ce travail nous a permis d'obtenir une base de données constituée de toutes les interactions et traces définies par notre protocole de recueil, issues de l'expérimentation Copéas. Le travail présenté dans cet article s'intéresse à la représentation et au traitement des données. Nous nous sommes orientés vers la recherche de patterns, i.e. suite d'actions récurrentes.

Ce travail s'inscrit dans le projet Mulce<sup>1</sup> et vise à proposer un outil de recherche de patterns dans un ensemble organisé de données issues d'interactions en ligne. Les patterns résultats sont ensuite présentés au chercheur, ils constituent une aide, une phase préalable à l'analyse, en lui permettant de visualiser les séquences, schèmes récurrents.

Dans une première partie nous présentons l'expérimentation qui nous a amené à travailler sur la recherche de patterns et les traces recueillies. Nous présentons ensuite les travaux sur la recherche de patterns puis l'implémentation de notre outil. Nous terminons par les tests effectués et les premiers résultats.

## 2. Le corpus Copéas

### 2.1. L'expérimentation

Le projet de recherche pluridisciplinaire Copéas (Communication Pédagogique en environnement orienté Audio Synchrone) mené par deux équipes (sciences du langage et informatique) a permis de réaliser une expérimentation écologique qui s'est déroulée sur 16 séances (8 par groupe) dans un environnement audio graphique synchrone. Il s'agit d'une formation qui vise à développer des compétences d'expression orale dans un contexte professionnel en anglais langue seconde chez 14 apprenants en master professionnel FOAD (Université de Franche-Comté). Le scénario de la formation, conçu par les tuteurs anglophones de l'Open University,

---

<sup>1</sup>Mulce : <http://mulce.univ-fcomte.fr>

propose des activités collaboratives sur la négociation de critères d'évaluation de sites Web pédagogiques.

La plateforme audio graphique synchrone utilisée dans cette expérimentation est Lyceum : plateforme développée et utilisée au sein de l'Open University (GB). En tant qu'environnement d'apprentissage audio graphique synchrone, Lyceum permet à un enseignant/tuteur de retrouver, à distance, des apprenants en mode synchrone. Les différents participants connectés à l'environnement peuvent donc se parler en temps réel, intervenir dans un chat et voir/modifier simultanément des productions textuelles ou graphiques.

L'interface de Lyceum rassemble trois composants activables simultanément :

- un composant spatial pour se déplacer du hall d'entrée dans les salles de travail ou techniques ;
- un composant rassemblant les outils de communication synchrone : module audio (avec outil de vote) et chat ;
- un composant permettant d'intégrer selon les besoins de l'activité différents modules de production collaborative : traitement de texte, carte conceptuelle et/ou tableau blanc.

Dans Lyceum, tous les acteurs (tuteur et apprenants) disposent de la même interface et des mêmes droits. Pour plus de précisions cf. [VETTER 04].

## **2.2. Recueil et organisation des données**

A l'issue de l'expérimentation, nous avons proposé une organisation des données d'interactions et des traces [BETBEDER et al. 06]. A l'inverse de certaines recherches (par exemple [AVOURIS et al. 04]), nous ne pouvions tracer automatiquement les interactions issues de la plateforme, nous avons donc décidé dans le protocole expérimental d'enregistrer les interactions dans la plateforme par vidéo. Par la suite les actions audio et les actions de production ont été transcrites. L'ensemble des interactions et traces sont donc stockées dans une base de données structurée. Chaque interaction (audio, textuelle, graphique ou iconique) ou trace (entrées/sorties) constitue l'un des 13000 enregistrements issus des 16 sessions. Les éléments caractéristiques principaux d'une action sont : un canal (audio, vote, chat, outil de production, etc.), une valeur précisant le contenu de l'action, une date de début et éventuellement une date de fin, l'acteur l'ayant réalisée et le lieu (salle virtuelle de travail) dans lequel elle s'est produite.

Cette organisation permet de retrouver une action précise, d'effectuer des calculs statistiques sur le nombre ou la durée des tours de parole des acteurs [VETTER & CHANIER 06] ou bien de lire toutes leurs interventions écrites et orales (transcrites) (pour une analyse qualitative sur le niveau d'anglais). Par contre ce travail n'est pas suffisant pour aider à l'interprétation des données notamment des analyses sur

l'enchaînement des actions, sur la multimodalité (organisation des groupes sur les canaux utilisés).

La difficulté d'interprétation des traces est actuellement un champ largement étudié (par exemple [HERAUD et al. 05], [AVOURIS et al. 05], [COURTIN & TALBOT 06]). Nous pensons que la recherche de patterns, appliquée à un grand volume d'interactions, peut faire ressortir des suites d'actions récurrentes qui peuvent aider le chercheur dans son analyse. Elle permettrait une analyse complémentaire aux traitements statistiques qui mettent à plat les différentes actions, sans considérer leur enchaînement. Ce qui nous importe ici, est de détecter si des habitudes ou des règles d'interaction s'établissent dans un groupe d'apprenants à l'intérieur d'un environnement multimodal de communication et de production. La recherche de patterns est-elle capable de repérer les habitudes ou stratégies des acteurs ? Permettrait-elle de repérer par exemple qui reprend la parole après un silence ? Qui initie une conversation ? Plus généralement, peut-on trouver des règles qui montrent comment les acteurs utilisent les différents modes de communication et de production pour leur travail collectif.

### 3. La recherche de patterns

Nous distinguons la *reconnaissance* de la *recherche* de patterns : La reconnaissance de patterns pré-suppose l'existence d'un modèle dont on va rechercher les occurrences parmi un ensemble de données. A l'inverse, la recherche de patterns (pattern mining) cherche à découvrir des motifs récurrents sans qu'aucun modèle ne soit suggéré. Cette technique est issue du domaine de la fouille de données ou *data mining*. Le but de notre outil est de calculer des séquences récurrentes dans une base de données des interactions.

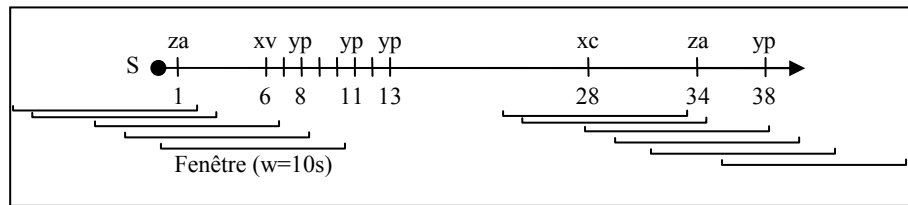
A l'origine, le *pattern mining* a été utilisé à des fins commerciales pour analyser « le panier de la ménagère » puis développer des stratégies marketing sur les achats alimentaires des consommateurs. A partir de l'analyse de nombreuses transactions, on a pu déterminer par exemple que dans 33% des cas, lorsqu'un panier contient de la bière, il contient également des chips [AGRAWAL et al. 93]. Dans ce cas, les éléments ne sont pas ordonnés (dans le panier) : seule l'appartenance d'un élément au panier est significative.

Pour les données totalement ordonnées (par exemple de type ADN) et une recherche de séquences parfaitement identiques (n'incluant aucun élément étranger), un algorithme a été proposé dans [AGRAWAL & SRIKANT 95].

Dans d'autres applications, comme l'analyse des séquences d'actions d'un utilisateur, ou dans la détection d'erreurs en cascade dans les réseaux de télécommunication, les événements sont temporellement situés [MANNILA et al. 97]. La prise en compte du facteur temps introduit deux notions essentielles. D'une part, la base de données source à analyser n'est plus considérée comme un ensemble, mais comme une séquence ordonnée, et d'autre part, la proximité entre deux

événements (dans le temps) pourra être interprétée comme une relative dépendance entre ces événements. En effet, on souhaite chercher si l'événement  $E_1$  est souvent suivi (d'assez près) d'un événement  $E_2$  sans pour autant que les deux événements soient nécessairement formellement contigus dans la séquence : des événements produits par une autre source ou un autre acteur peuvent s'insérer de façon chaotique entre  $E_1$  et  $E_2$ .

Formellement, nous considérons donc une séquence  $S$  d'entrée contenant une liste d'événements temporellement situés. Soit  $A$  nommé alphabet, l'ensemble des types d'événements possibles, chacun des événements de  $S$  sera représenté par un couple  $(E, t)$  tel que  $E \in A$  et  $t \in \mathbb{N}$ , où  $t$  représente la date de début de l'événement.



**Figure 1.** Exemple d'une séquence d'événements

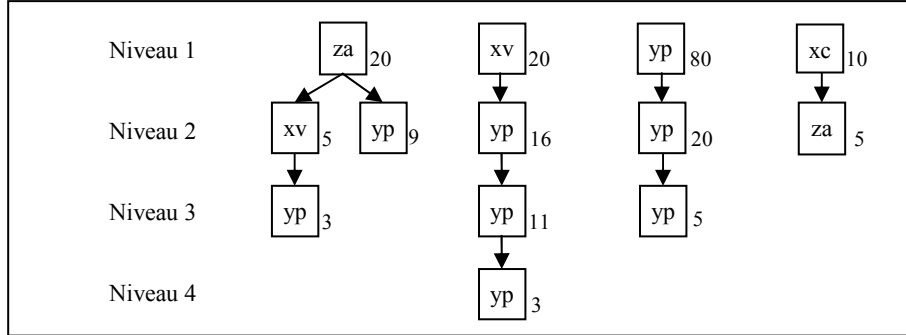
Exemple : Considérons 3 acteurs  $x$ ,  $y$  et  $z$  et 4 canaux : audio ( $a$ ), chat ( $c$ ), vote ( $v$ ) et production ( $p$ ). Si nous souhaitons différencier chaque action par l'acteur et le canal utilisé, notre alphabet représentant l'ensemble des types d'actions possibles sera donc :  $A = \{xa, xc, xv, xp, ya, yc, yv, yp, za, zc, zv, zp\}$ . Soit  $S$ , une séquence commençant par un tour de parole (audio) de  $z$  à  $t_1=1s$  et finissant par une action de production de  $y$  à  $t_8=37s$ . Une telle séquence  $S$  est illustrée sur la figure 1 et définie par :

$$S = \{(za,1),(xv,6),(yp,8),(yp,11),(yp,13),(xc,28),(za,34),(yp,38)\}$$

### 3.1. Calcul des motifs fréquents

La recherche de motifs fréquents permet de dégager des séquences ou épisodes d'actions dont le nombre élevé d'occurrences peut être porteur d'informations. Ces informations pourront être soit exploitées directement soit soumises à d'autres traitements afin d'obtenir des informations plus précises (avec le calcul de règles d'association, par exemple).

Un motif est une séquence ordonnée de types d'événements, notée  $M=(E_1.E_2...E_n)$  où  $E_i \in A$  et où  $n \in \mathbb{N}$ , représente la longueur de la séquence aussi appelée « niveau » du motif. L'algorithme Winepi [MANNILA et al. 97] fait glisser une fenêtre de largeur  $w$  le long de la séquence d'entrée. Pour chaque position de la fenêtre, on identifie tous les motifs. A l'issue de cette recherche sur les motifs, faite de manière incrémentale, on obtient pour chaque motif, le nombre de fenêtres dans lesquelles il a été identifié.



**Figure 2.** Arborescence des motifs repérés sur la séquence exemple *S*

La figure 2 donne, sous forme arborescente, la liste de tous les motifs qui apparaissent dans les fenêtres de l'algorithme Winepi appliqué à la séquence exemple et pour une fenêtre de largeur  $w=10$ s. Nous constatons que tous les motifs sont de niveau inférieur ou égal à 4. L'arbre le plus à gauche dont la racine est « za » nous donne les indications suivantes : le motif (za) a été identifié dans 20 fenêtres, le motif (za.xv) dans 5, (za.xv.yp) dans 3 et (za.yp) dans 9 fenêtres.

Mais dans Winepi, ne sont conservés que les motifs dont la fréquence dépasse une fréquence seuil  $f_{\min}$  donnée en paramètre. Soient  $t_1$  la date la plus petite de la séquence *S* et  $t_n$  la plus grande,  $w$  la taille de la fenêtre. Considérons  $Nw_j$  le nombre de fenêtres dans lesquelles apparaît le motif  $M_j$ , alors la fréquence  $f_j$  du motif  $M_j$  est donnée par :

$$f_j = \frac{Nw_j}{(t_n - t_1 + w - 1)}$$

Bien que cette notion soit nommée « fréquence » dans [MANNILA et al. 97], elle n'est pas nécessairement comprise entre 0 et 1. Dans notre exemple, les  $M_j$  (reportés à droite de la feuille de chaque motif sur la figure 2) seront donc divisés par 46. La fréquence la plus basse est 0,065, obtenue pour les motifs (za.xv.yp) et (xv.yp.yp.yp), tandis que la plus haute est de 1,739 pour le motif (yp). Ces valeurs sont relativement élevées car sur l'exemple, l'amplitude de la séquence est faible.

Les fréquences des motifs de niveau 1 ne nous donnent pas plus d'information qu'une simple statistique des occurrences de chaque type d'événement. En revanche, les fréquences des motifs plus grands nous renseignent sur la fréquence de successions d'événements non nécessairement contigus, mais toujours survenus dans un laps de temps restreint (inférieur ou égal à la taille  $w$  de la fenêtre). Si une telle succession d'événements est fréquente, nous suspectons une relation de dépendance (de cause à effet) entre les événements qui la composent. C'est justement ce que nous propose le calcul des règles d'association.

### 3.2. Les règles d'association

Prenons un motif  $M$  composé de deux parties  $M_1$  et  $M_2$  tel que  $M=M_1.M_2$ . Si l'arborescence des motifs fréquents nous montre que le motif  $M_1$  est (souvent) suivi du motif  $M_2$ , nous l'interprétons comme le fait que  $M_2$  est une conséquence de  $M_1$ .  $M$  est alors appelé règle,  $M_1$  sa prémisse et  $M_2$  sa conclusion. Le degré de confiance de cette règle dépend de la fréquence du motif  $M$ , de celle de  $M_1$ , mais aussi et surtout du fait que  $M_2$  est ou non le seul motif qui puisse succéder à  $M_1$ , et s'il y en a d'autres, la proportion d'apparition de  $M_2$  par rapport aux autres  $M_k$ . A partir de l'arbre des motifs fréquents, si l'on considère la prémisse  $M_1$  de fréquence  $f_1$ , et toutes les conclusions  $M_k$ , de fréquence  $f_k$ , descendantes de  $M_1$ , alors toutes les règles  $R_k=M_1.M_k$  auront pour confiance  $C_{1k}=f_k/f_1$ .

Dans l'algorithme implanté, nous ne sélectionnons que les règles dont la confiance dépasse une valeur seuil  $C_{min}$ .

### 3.3. Recherche de patterns séquentiels

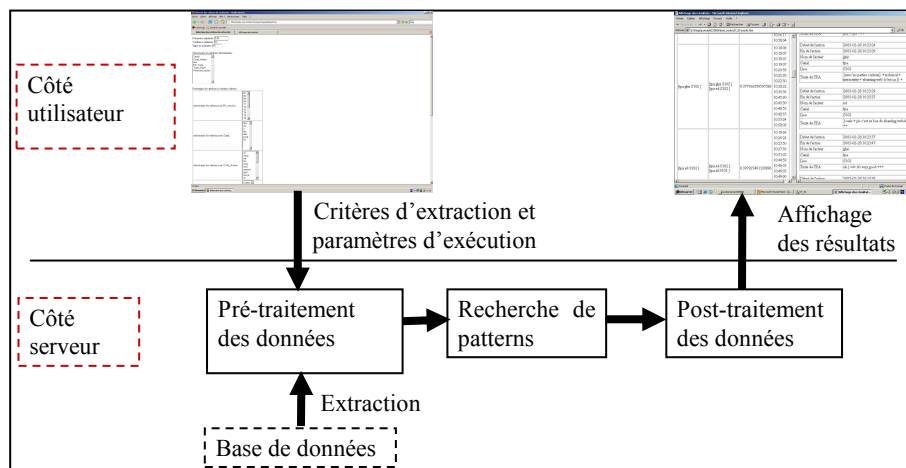
Si la recherche de patterns est peu utilisée dans le domaine des EIAH, nous pouvons cependant citer les travaux de [MERCERON & YACEF 05] et [CHEYPE 06]. Les premières travaillent sur les résultats d'exercices. Elles repèrent au moyen des patterns les étudiants « à risque » (risque de ne pas réussir un exercice si d'autres n'ont pas été réussis) et des comportements d'apprenants. Les travaux de Cheype concernent des traces hétérogènes issues d'activités collectives. L'objectif est de détecter des indices sur les suites d'actions menant à l'échec ou la réussite d'exercices. Dans les deux cas la recherche de patterns est utilisée pour aider l'enseignant, pour avoir une meilleure appréciation du travail des étudiants.

## 4. Implémentation

Nous avons choisi d'implémenter l'algorithme Winepi [MANNILA & al. 97], son principe consiste à calculer d'abord l'ensemble des motifs fréquents puis à déterminer quelles règles d'association dépassent le seuil de confiance  $C_{min}$  fixé par l'utilisateur. Le calcul des motifs fréquents est guidé par deux paramètres essentiels : la fréquence minimale  $f_{min}$  et la taille  $w$  de la fenêtre. La fréquence minimale détermine le seuil au-delà duquel un motif est considéré comme fréquent. La taille de la fenêtre fixe l'intervalle de temps dans lequel doit être compris un motif.

L'algorithme de recherche de patterns ne traite pas directement avec les données : une action n'est pas définie par ses attributs (par exemple code acteur, date, canal), mais par un identifiant numérique caractérisant le type d'action. Une phase de pré-traitement (cf. Figure 3) permet donc d'attribuer à chaque type d'action (e.g. [canal, acteur]) un identifiant unique et de l'affecter à chaque action du même type dans la séquence (et réciproquement en phase de post-traitement).





**Figure 3.** Implémentation de l'outil de recherche de patterns

#### 4.1. Critères d'extraction et paramètres d'exécution

Le paramétrage possible concerne les critères d'extraction et les paramètres d'exécution de l'algorithme. Pour cela, il est possible de « jouer » sur trois paramètres généraux qui ont un impact sur la recherche de patterns. Il s'agit :

- du seuil minimal de fréquence des motifs ;
- du seuil de confiance minimum pour les règles d'association, qui fixe un degré de confiance minimal compris entre 0 et 1 pour les règles produites ;
- de la taille de la fenêtre qui fixe la taille maximale des motifs en secondes.

Pour la recherche des motifs, il est nécessaire de fixer les paramètres de différenciation des actions. L'utilisateur doit définir un type d'action ; pour cela il choisit parmi les attributs qui caractérisent une action, lesquels sont discriminants. Dans nos exemples nous avons choisi le canal et le code acteur. Deux actions sont considérées comme étant de même type si elles ont le même canal et le même acteur (par exemple [audio, tim]). L'utilisateur peut également filtrer les actions, en ne considérant par exemple que celles produites au cours d'une session donnée, dans un lieu donné.

#### 4.2. Affichage des résultats

La liste des règles<sup>2</sup> résultats est affichée (prémisse et conclusion), ainsi que le degré de confiance et les instances de celle-ci. Les dates des instances correspondent

<sup>2</sup> Nous utilisons indifféremment les termes « règle » et « pattern ».

à l'heure du début de la fenêtre dans laquelle se déroule la règle. La liste complète des actions est également affichée.

## 5. Tests et résultats

### 5.1. Tests

A l'issue de ce développement, nous avons effectué des tests de performance sur le temps de calcul, le nombre de patterns trouvés, le nombre de types d'actions. Les tests préliminaires nous ont permis de déterminer les valeurs tests des paramètres [TISSOT 06] : fenêtre de 60 secondes, seuil minimal de fréquence des motifs à 0.05 et seuil de confiance minimum à 0.3. Pour les critères d'extraction nous avons choisi comme attributs discriminants les canaux et les acteurs, et la restriction s'effectue par session (T1 à T8) et par salle (Lobby, 101 et 102). Lors des premières séances, seule la salle 101 a été utilisée tandis que pour les sessions T5, T7 et T8 le groupe s'est divisé en deux sous groupes répartis en deux salles 101 et 102. Les résultats ici de la salle Lobby ne sont par pris en compte car peu intéressants, cette salle est principalement utilisée en début de session (passage obligé lors de la connexion à la plateforme) pour les salutations (elle ne dispose pas des outils de production). D'un point de vue quantitatif il s'avère que le nombre de motifs et de règles ne sont pas proportionnels au nombre d'actions. Selon les sessions et lieux le nombre de règles varient entre 7 et 48. Le nombre de motifs et le nombre de règles sont relativement proportionnels.

### 5.2. Résultats

L'étude des premiers résultats nous permet d'ores et déjà de présenter quelques conclusions :

- Avec ces valeurs des paramètres il existe peu de patterns informatifs (cf. discussion) de plus de 3 niveaux, i.e. des séquences de plus de trois actions successives.
- Nous observons dans la grande majorité des résultats des patterns uni-modaux (75,6%) de forme (1) ou (2). Ceci porte à croire que les acteurs utilisent le même canal pour interagir avec un autre acteur.

(1) [audio, x] => [audio, y] ou [audio, x], [audio, y] => [audio, z]

(2) [prod, x] => [prod, y] ou [prod, x], [prod, y] => [prod, x]

Des exceptions sont cependant à noter : dans deux des sessions, le nombre de patterns bi-modaux (audio et production) est significativement plus important. Ces sessions correspondent au travail du même sous groupe. Ces règles montrent les interactions entre trois acteurs qui alternent entre audio et production.

[audio, al], [prod, chris] => [audio, her]

[audio, her] => [prod, her]

- Le rôle du tuteur ressort fortement grâce aux patterns. Les calculs statistiques sur la base de données nous avaient déjà permis de voir sa participation en terme de nombre de tours de parole et de temps de parole (52% du temps parlé, i.e excluant les silences). De même que dans toutes les sessions on observe des patterns de la forme (1) correspondant à la difficulté des apprenants à répondre immédiatement au tuteur, on observe beaucoup de patterns de la forme (2) qui indiquent des actions d'« explicitation » du tuteur.

(1) [audio, tuteur] => [audio, silence]

(2) [audio, tuteur], [audio, silence<sup>3</sup>] => [audio, tuteur]

- Alors que la proportion d'actions de chat est significative (12,2%), on observe très peu de patterns (5%) faisant apparaître un acte de chat. Ceci montre que le chat n'est pas utilisé en mode unique. Par contre, les patterns ne permettent pas de faire ressortir les interactions de chat effectuées comme soutien à l'audio sans en interrompre les interventions audio et les discussions parallèles (actes sporadiques).

- A l'inverse, des patterns sur les votes émergent bien que ces actions soient statistiquement peu représentées (4.7%). Ils correspondent aux votes des apprenants sur les questions de vérification de compréhension des consignes du tuteur.

[vote, x] => [vote, y]

[vote, x] => [audio, tuteur]

## 6. Discussion et conclusion

Cet outil propose des résultats devant être interprétés par le chercheur. Un certain nombre de règles trouvées sont peu intéressantes. Par exemple, des patterns proposent une succession d'actions de même type et même acteur (de niveau allant jusqu'à 6). Ces patterns signifient que pendant un certain laps de temps un acteur a travaillé seul.

Une question se pose sur la pertinence de l'utilisation des règles. Nous nous demandons si dans le cas d'apprentissage il est indispensable de calculer les règles. Celles-ci permettent d'introduire un lien de causalité entre des suites d'actions. Cependant, le calcul des motifs les plus fréquents (de tout façon préalable) peut-il être également un résultat à afficher et à analyser en soi.

Cet outil fonctionnel permet d'ores et déjà de révéler des patterns intéressants pour l'analyse de la formation. Par exemple, l'outil nous a permis de révéler le pattern concernant le rôle du tuteur ainsi que celui du sous-groupe des trois acteurs dont nous avons eu l'intuition en suivant toutes les sessions.

---

<sup>3</sup> Dans les codes de transcription définis nous avons choisi de coder les silences comme un tour de parole audio afin de pouvoir les identifier.

Les actions de production du deuxième groupe vont prochainement être transcrites, nous pourrons ainsi comparer les patterns trouvés dans les deux groupes, notamment pour étudier l'intervention du tuteur et voir si l'on retrouve des patterns « d'explicitation » dans le groupe de meilleur niveau.

Des évolutions sont envisagées. Actuellement les algorithmes de recherche de patterns traitent des données ponctuelles, or certaines données que nous étudions sont caractérisées par une durée. Nous avons fait le choix de ne prendre en compte pour ces actions que la date de début de l'action. De plus, il serait intéressant de pouvoir utiliser des variables pour décrire les patterns. Cette utilisation nous permettrait de trouver des patterns de la forme un acteur utilise l'audio => un autre modifie la production ou encore le tuteur parle => un apprenant prend la parole.

Ce travail s'inscrit dans le projet Mulce qui vise à proposer un environnement d'échange et de mise à disposition de corpus d'apprentissage. Pour cela, nous travaillons [NORAS 06] à définir une spécification pour décrire un corpus d'apprentissage (incluant scénario pédagogique, traces, données, etc.). S'appuyant sur IMS-CP, et IMS-LD, cette spécification intégrera les différents éléments constitutifs d'un corpus : les traces - interactions - productions, le contexte de recherche, le scénario pédagogique, l'instanciation et les observations du contexte.

Aujourd'hui, l'outil de recherche de patterns utilise en entrée une base de données spécifique à l'expérimentation Copéas (session, lieu, canal, etc.). Courant 2007 l'outil sera mis à disposition, le prototype présenté dans cet article sera redéveloppé afin de pouvoir intégrer des données plus génériques mises au format de la spécification. Les chercheurs pourront ainsi déposer un corpus préalablement mis au format de la spécification puis utiliser l'outil de recherche de patterns à des fins d'analyse.

## Remerciements

Mulce (Échange de corpus d'apprentissage multimodaux) est un projet soutenu par l'Agence nationale de la Recherche (ANR-06-CORP-006) dans le cadre du programme "Corpus et Outils de la Recherche en Sciences Humaines et Sociales". Il rassemble des équipes des laboratoires LASELDI et LIFC (Université de Franche-Comté), CREET (The Open University) et LIP6 (Université Paris 6), coordonnées respectivement par Thierry Chanier, Christophe Reffay, Marie-Noelle Lamy et Jean-Gabriel Ganascia.

## 7. Bibliographie

[AGRAWAL et al. 93] Agrawal, R., Imielinski, T., Swami, A., « Mining association rules between sets of items in large databases », *Actes de ACM SIGMOD Conference on Management of Data (SIGMOD '93)*, Washington D.C., 1993, p. 207–216.

- [AGRAWAL & SRIKANT 95] Agrawal, R. Srikant, R., « Mining sequential patterns », *Actes de the Eleventh International Conference on Data Engineering (ICDE'95)*, Taipei, Taiwan, 1995, p. 3-14.
- [AVOURIS et al. 05] Avouris, N., Komis, V. Fiotakis, G., Margaritis, M., Voyiatzaki, E., « Logging of fingertip actions is not enough for analysis of learning activities », *Actes du Workshop Usage Analysis in learning systems, AIED 2005*, Amsterdam, Pays-Bas, juillet 2005.
- [AVOURIS et al. 04] Avouris, N., Komis, V., Fiotakis, G., Dimitracopoulou, A., Margaritis, M., « Method and Tools for analysis of collaborative problem-solving activities », *Actes de ATIT2004, First International Workshop on Activity Theory Based Practical Methods for IT Design*, Copenhagen, Danemark, septembre 2004, p. 5-16.
- [BETBEDER et al. 06] Betbeder, M.-L., Reffay, C., Chanier, T., « Environnement audio-graphique synchrone : recueil et transcription pour l'analyse des interactions multimodales », *Actes de JOCAIR'2006*, Amiens, 6-7 juillet 2006, p. 406-420.
- [CHEYPE 06] Cheype, A., « Recherche de motifs séquentiels pour guider l'interprétation des traces d'apprentissage dans un EIAH ». *Actes de RJC-EIAH'2006*, Evry, 11-12 mai 2006, p.123-130.
- [COURTIN & TALBOT 05] Courtin, C., Talbot, S., « An Architecture to Record Traces in Instrumented Collaborative Learning Environments », *Actes IADIS de CELDA 2005*, Porto, Portugal, 14-16 décembre 2005, p.301-308
- [HERAUD et al. 05] Heraud, J.-M., Marty, H.-C., France, L., Carron, T., « Une aide à l'interprétation de traces : application à l'amélioration de scénarios pédagogiques », *Actes de EIAH 2005*, Montpellier, mai 2005, p. 237-248.
- [MANNILA et al. 97] Mannila, M., Toivonen, H., Verkamo, A., *Discovery of Frequent Episodes in Event Sequences*,. Springer Science+Business Media B.V., Formerly Kluwer Academic Publishers B.V., 1997, p 259 – 289.
- [MERCERON & YACEF 05] Merceron, A., Yacef, K., *Educational data mining: a case study*, IOS Press, 2005.
- [NORAS 06] Noras, M., « Un besoin de spécification des corpus de formation en ligne », *Actes de RJC-EIAH'2006*, Evry, 11-12 mai 2006, p. 165-172.
- [SETTOUTI et al. 06] Settouti, L.-S., Prié, Y. Mille, A. Marty, J.-C., « Systèmes à base de traces pour l'apprentissage humain », *Actes de TICE 2006*, Toulouse, octobre 2006.
- [TISSOT 06] Tissot, R., Recherche de patterns dans un corpus d'actions multimodales, Mémoire de Master informatique, Université de Franche-Comté, août 2006, 60p.
- [VETTER & CHANIER 06] Vetter, A., Chanier, T., « Supporting oral production for professional purposes in synchronous communication with heterogeneous learners », *ReCALL*, vol.18, n°1, 2006, p. 5-23.
- [VETTER 04] Vetter, A., « Les spécificités du tutorat à distance à l'Open University : enseigner les langues avec Lyceum », *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, vol. 7, 2004, p. 107-129.